Multi-task Representation Learning for Demographic Prediction

Pengfei Wang[1], Jiafeng Guo[2], Yanyan Lan[2], JunXu[2], and Xueqi Cheng[2]

Key Laboratory of Network Data Science and Technology,CAS,
Institute of Computing Technology, Chinese Academy of Sciences
wangpengfei@software.ict.ac.cn[1]
{guojiafeng,lanyanyan,junxu,cxq}@ict.ac.cn[2]

Abstract. Demographic attributes are important resources for market analysis, which are widely used to characterize different types of users. However, such signals are only available for a small fraction of users due to the difficulty in manual collection process by retailers. Most previous work on this problem explores different types of features and usually predicts different attributes independently. However, manually defined features require professional knowledge and often suffer from under specification. Meanwhile, modeling the tasks separately may lose the ability to leverage the correlations among different attributes. In this paper, we propose a novel Multi-task Representation Learning (MTRL) model to predict users' demographic attributes. Comparing with the previous methods, our model conveys the following merits: 1)By using a multi-task approach to learn the tasks, our model leverages the large amounts of cross-task data, which is helpful to the task with limited data; 2)MTRL uses a supervised way to learn the shared semantic representation across multiple tasks, thus it can obtain a more general and robust representation by considering the constraints among tasks. Experiments are conducted on a real-world retail dataset where three attributes (gender, marital status, and education background) are predicted. The empirical results show that our MTRL model can improve the performance significantly compared with the state-of-the-art baselines.

Keywords: Multi-task, demographic prediction, representation learning

1 Introduction

Acquiring users' demographic attributes is crucial for retailers to conduct market basket analysis [18], adjust marketing strategy [9], and provide personalized recommendations [20]. However, in practice, it is difficult to obtain users' demographic attributes, because most users are reluctant to offer their detailed information or even refuse to give their demographics due to privacy and other reasons. This is particularly true for traditional offline retailers[1], who collect

[1] In our work, we mainly focus on traditional retailers in offline business rather than those in online e-commerce, where no additional behavioral data rather than trans-

users' demographic information mostly in a manual way (e.g. requiring costumers to provide demographic information when registering some shopping cards).

In this paper, we try to inference users' demographic attributes based on users' purchase history. Although some recent studies suggest that demographic attributes are predictable from different behavioral data, such as linguistics writing [5], web browsing [16], electronic communications [8, 12] and social media [14, 23], to our best knowledge, seldom practice has been conducted on purchase behaviors in the retail scenario.

The previous work about demographic prediction usually predicts demographic attributes separately based on manually defined features [3, 17, 19, 22, 23]. For example, Zhong et al. [23] predicted six demographic attributes (i.e., gender, age, education background, sexual orientation, marital status, blood type and zodiac sign) separately by merging spatial, temporal and location knowledge features into a continuous space. Obviously, manually defined features usually require professional knowledge and often suffer from under specification. Meanwhile, by taking each attribute as independent prediction task, some attributes may difficult to predict due to the insufficient data in training. Some recent studies proposed to take the relations between multiple attributes into account [3, 22]. For example, Dong et al. [3] employed a Double Dependent-Variable Factor Graph model to predict gender and age simultaneously. Zhong et al. [22] attempted to capture pairwise relations between different tasks when predicting six demographic attributes from mobile data. However, these methods still rely on various human-defined features which are often costly to obtain.

To tackle the above problem, in this paper we propose a Multi-task Representation Learning(MTRL) model is used to predict users' gender, martial status, and education background based on users' purchase history. MTRL learns shared semantic representations across multiple tasks, which benefits from a more general representation for prediction. Specifically, we characterize each user by his/her purchase history using the bag-of-item representations. We then map all users' representations into semantic space learned by a multi-task approach. Thus we can obtain a more general shared representation to guide the prediction task separately. Compared with previous methods, the major contributions of our work are as follows:

- We make the first attempt to investigate the prediction power of users' purchase data for demographic prediction in the retail scenario.
- We apply a multi-task learning framework (MTRL) for our problem, which can learn a shared robust representation across tasks and alleviate the data sparse problem.
- We conduct extensive experiments on a real-world retail dataset to demonstrate the effectiveness of the proposed MTRL model as compared with different baseline methods.

The rest of the paper is organized as follows. After a summary of related work in Section 2, we describe the problem formalization of demographic prediction

actions are available for analysis. Hereafter we will use retail/retailer for simplicity when there is no ambiguity.

in the retail scenario in Section 3. In section 4 we present our proposed model in detail. Section 5 concludes this paper and gives the future work.

2 Related Work

In this section we briefly review three research areas related to our work: demographic attribute prediction, multi-task approach, and representation learning.

2.1 Demographic Attribute Prediction

Demographic inference has been studied in different scenarios for more than fifty years. Early stage work on demographic prediction attempted to predict demographic attributes based on the linguistics writing and speaking. For example, Schler et al. [19] found that there are significant differences in both writing style and content between male and female bloggers as well as among authors of different ages. Otterbacher [17] used a logistic regression model to infer users' gender based on content of reviews.

Furthermore, researchers use internet information to predict demographic attributes [8, 16]. For example, Torres [4] found a clear relation between the reading level of clicked pages and demographic attributes such as age and education background. Hu et al. [8] calculated demographic tendency of web pages, and modeled users' demographic attributes through a discriminative model. In [1], Bi et al. propose to infer the demographic attributes of search users based on the models trained on the independent social datasets. They demonstrated that by leveraging social and search data in a common representation, they can achieve better accuracy in demographic prediction.

Additionally, the fast development of online social networks and mobile computing technologies bring a new opportunity to identify users' demographic attributes. Mislove [14] found that users with common profiles were more likely to be friends and often formed a dense community. Zhong et al. [22] proposed a supervised learning framework to predict users' demographic attributes based on mobile data. Dong et al. [3] focused on micro-level analysis of the mobile networks to infer users' demographic attributes. Culotta et al. [2] fitted a regression model to predict users' demographic attributes using information on followers of each website on Twitter.

As we can see, most existing work on demographic prediction focused on designing different features for the prediction tasks. Besides, to the best of our knowledge, seldom practice has been conducted on demographic prediction based on purchase behaviors in the retail scenario.

2.2 Multi-task Approach

The advantage of multi-task approach is to improve the generalization performance by leveraging the information contained in the related tasks. A typical way of multi-task approach is to learn tasks in parallel with a shared representation [3, 21, 22]. Many algorithms have been proposed to solve multi-task learning tasks. For example, Micchelli et al. [13] discussed how various kernels can be used to model relations between tasks and presented linear multi-task learning algorithms. Evgeniou et al. [6] presented an approach to multi-task learning based on the minimization of regularization functionals.

2.3 Representation Learning

Learning representations of the data makes it easier to extract useful information when building classifiers or other predictors, without extracting features in a manual way. That is the reason why representation learning has attracted more and more attention and becomes a field in itself in the machine learning community.

Recently, plenty remarkable successes have been achieved based on representation learning in various applications in both academia and industry. For example, Alex Graves et al. [7] designed a deep recurrent neural network for speech recognition and obtained the best score on an evaluation benchmark. Krizhevsky et al. [10] proposed to use convolutional neural network to classify images, achieving record-breaking results. Mnih [15] proposed three graphical models to define the probability of observing next word in a sequence, leveraging distributed representations.

In this work, we propose to use the multi-task approach to learn a shared representation for demographic prediction in the retail scenario, a new application area where representation learning might be helpful, especially to the task with limited data.

3 Our Approach

In this section, we first give the motivation of our work, then we introduce the formalization of demographic prediction problem in the retail scenario. After that, we describe the proposed MTRL in detail. Finally, we present the learning procedure of MTRL.

3.1 Motivation

Obviously, a fundamental problem in demographic prediction based on users' behavioral data is how to represent users. Many existing work investigated different types of human defined features [3, 17, 22]. However, defining features manually costs time since expertise knowledge is required and one has to do the same process repeatedly. Moreover, human defined features may often suffer from under specification since it is difficult to identify those hidden complicated factors for prediction tasks. Recent work mainly employs unsupervised feature learning methods [8, 12, 23], like Singular Vector Decomposition (SVD), to automatically extract low-dimension features from the raw data. However, the features learned

Table 1: List of demographic attributes

Demographic Attributes	Values
gender	male, female
marital status	single, married
education background	doctor, master, bachelor, college, high school, middle school

in an unsupervised manner may not be suitable for the prediction tasks. Therefore, concerning the weakness of extracting features humanly, in this paper we proposed to automatically learn representations of users for demographic prediction through a supervised method. Furthermore, some attributes are difficult to obtain(for example, only 8.96% of users offer their education background in the BeiRen dataset we used). Thus the sparseness of data aggravates the difficulty of modeling the task separately [2, 12, 23]. In addition, modeling the tasks independently may ignore the correlations among these attributes.

Motivated by all these issues, inspired by [11], in this paper we propose a multi-task approach to learn a general representation to predict users' demographic attributes.

3.2 Problem Formalization

In this work, we try to predict multiple demographic attributes given users' behavioral data in the retail scenario. Specifically, each user can be characterized by his/her purchase history, i.e., a set of items. The demographic attributes we are interested include gender, marital status, and education ground, which are valuable signals for market analysis. The values of each attribute take are shown in Figure 1. Given a user, based on his/her purchase history, we want to predict all the unknown attributes.

Specifically, let $T = \{t_1, t_2, \ldots, t_n\}$ be a set of demographic prediction tasks (i.e., predicting demographic attributes). Let U be a set of users. Suppose the training set is composed of M instances, i.e.,

$$\{(x_{(1)}, y_{(1)}), (x_{(2)}, y_{(2)}), \ldots, (x_{(M)}, y_{(M)})\}$$

where $x_{(i)} \in X$ is a d-dimensional feature vector, representing the input of i-th user, and $y_{(i)}$ is the set of attribute labels of the i-th user. Note here $y^t_{(i)}$ denotes all the attribute labels under the t-th task $t \in T$ for the i-th user.

Based on the notations defined above, we try to learn a function to predict the unknown demographic attributes.

3.3 Multi-task Representation Learning Model

In this section, we now present the proposed MTRL model in detail. The feed-forward MTRL is shown in Figure 1. In the retail scenario, each user is characterized by his/her purchase history, i.e., a set of items. In MTRL, we take the

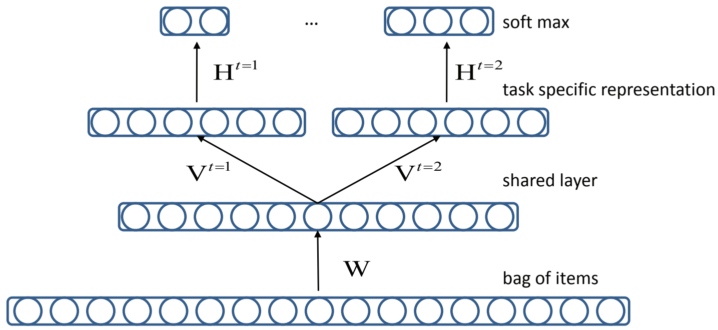


Fig. 1: The structure of Multi-task Representation Learning (MTRL) model. The lower two layers are shared across all the tasks, while top layers are task-specific. The input is represented as a bag of items. Then a non-linear projection **W** is used to generate a shared representation. Finally, for each task, additional non-linear projection **V** generates task-specific representations.

bag-of-item representation as the user input $x_{(i)}$, then the shared layer is fully connected to the input layer with weights Matrix $\mathbf{W} = [w_{h,s}]$:

$$\mathcal{Y}_{(i),s} = f(\sum_h w_{h,s} \cdot x_{(i),h})$$

where matrix $\mathbf{W}$ is responsible for generating the cross-task representation, $\mathcal{Y}_{(i),s}$ is the value of s-th node on the shared layer, $x_{(i),h}$ is the h-th value of $x_{(i)}$, and $f(z)$ is the logistic nonlinear activation:

$$f(z) = \frac{1}{1 + e^{-z}}$$

Based on the shared layer, for each task t, we use a transformation $V^t = [v_{s,j}]$ to map the shared representation into the task-specific representation by:

$$\mathcal{Y}^t_{(i),j} = f(\sum_s v^t_{s,j} \cdot \mathcal{Y}_{(i),s})$$

where t denotes the different tasks(gender, marital status, and education background), and $\mathcal{Y}^t_{(i),j}$ is the value of j-th node corresponding to the specific representation layer of task t.

After these, we use a softmax activation function to calculate the value of k-th node in the output layer:

$$\mathcal{Y}^t_{(i),k} = \frac{exp(\sum_j h^t_{j,k} \cdot \mathcal{Y}^t_{(i),j})}{\sum exp(\sum_j h^t_{j,k} \cdot \mathcal{Y}^t_{(i),j})}$$

where $H^t = [h^t_{j,k}]$ is the matrix that maps task specific representation to the output layer for task t, the k-th node in the output layer corresponds the value of the k-th label in task t.

Algorithm 1 Algorithm for Multi-task Learning Representation Model

```
Initialize model Θ: {W, V^t, H^t} randomly
iter=0
repeat
    iter ← iter + 1;
    for i=1,...,M do
        select a task t randomly for instance x_(i)
        compute the gradient∇(Θ)
        update model Θ ← Θ + ε∇(Θ)
    end for
until converge or iter > num
return W, V, H
```

The objective function of MTRL is then defined as the cross-entropy over the outputs of all the users and all tasks:

$$\ell_{MTRL} = \sum_t \sum_i \sum_k d^t_{(i),k} \ln \mathcal{Y}^t_{(i),k} + (1 - d^t_{(i),k}) \ln(1 - \mathcal{Y}^t_{(i),k}) - \lambda \|\Theta\|^2_F \tag{1}$$

where $d^t_{(i),k}$ is the real value of k-th node for user i under the task t, for example, for task t, if user i choose the k-th label, then $d^t_{(i),k} = 1$, else $d^t_{(i),k}$ equals 0. λ is the regularization constant and Θ are the model parameters (i.e. $\Theta = \{\mathbf{W}, \mathbf{V}^t, \mathbf{H}^t\}$).

3.4 Learning and Prediction

In order to learn parameters of MTRL model, we use the stochastic gradient decent algorithm, as shown in Algorithm 1. For each iteration, a task t is randomly selected, and parameters of the model is updated according to the task-specific objective.

With the learned parameters, the demographic prediction task is as follows. For each demographic task, we calculate the value of output layer through MTRL, and then the output node with the largest value is regarded as the label to the given user.

4 Experiments

In this section, we conduct empirical experiments to demonstrate the effectiveness of our proposed MTRL model on demographic attribute prediction in the retail scenario. We first introduce the experimental settings. Then we compare our MTRL model with the baseline methods to demonstrate the effectiveness of predicting users' demographic attributes in the retail scenario.

4.1 Experimental Settings

In this section, we introduce the experimental settings including the dataset, baseline methods, and evaluation metrics.

Dataset We conduct our empirical experiments over a real world large scale retail dataset, named BeiRen dataset. This dataset comes from a large retailer[2] in China, which records its supermarket purchase histories during the period from 2012 to 2013. For research purpose, the dataset has been anonymized with all the users and items denoted by randomly assigned IDs for the privacy issue.

We first conduct some pre-process on the BeiRen dataset. We randomly collected 100000 users. We extract all the transactions related to these users to form their purchase histories, then we remove all the items bought by less than 10 times and the users with no labels. After pre-processing, the dataset contains 64097 distinct items and 80540 distinct users with at least on demographic attribute. In average, each user has bought about 225.5 distinct items.

Baseline Methods We evaluate our model by comparing with several state-of-the-art methods on demographic attribute prediction:

- BoI-Single: Each user is represented by the items he/she has purchased with the Bag-of-Item representation and a logistic model [3] is learned to predict each demographic attribute separately.
- SVD-single: A singular value decomposition (SVD)[4] is first conducted over the user-item matrix to obtain low dimensional representations of users. Then a logistic model is learned over the low dimensional representation to predict each demographic attribute separately. This method has been widely used in the demographic attribute prediction task [8, 16, 23].
- SL: The Single Representation Learning model, which is a special case of MTRL when there is only one single task to learn. SL model has the same neural structure comparing with MTRL, just without considering the relationships among tasks.

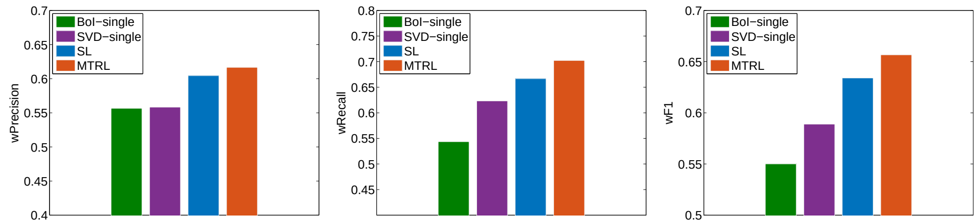
For SVD-single method, we run several times with random initialization by setting the dimensionality as 200. For MTRL and SL, we set the dimensionality of shared representation layer and task-specific representation layer as 200 and 100 respectively. The parameters are initialized with uniform distribution in the range $(-\sqrt{6/(fan_{in}+fan_{out})}, \sqrt{6/(fan_{in}+fan_{out})})$.

For each experiment, we run several times, we then compare the best results of different methods and demonstrate the results in the following sections.

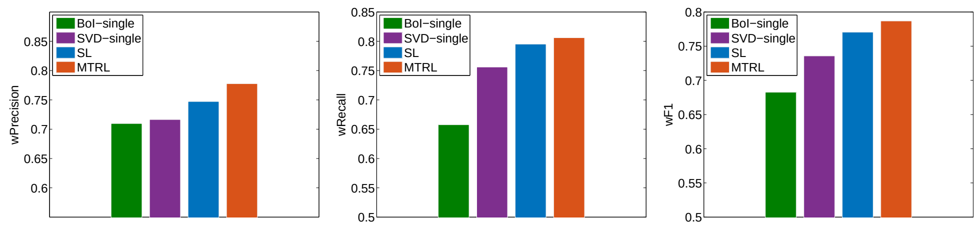
Evaluation Metrics we follow the idea in [3] to use weighted F1 as an evaluation metric. For task t, the weighted F1 is computed as follows:

$$\text{wPrecision} = \sum_{y \in t} \frac{\sum_i I(y^t_{(i)} = \mathcal{Y}^t_{(i)} \& y^t_{(i)} = y)}{\sum_i I(y^t_{(i)} = y)} \cdot \frac{\sum_i I(\mathcal{Y}^t_{(i)} = y)}{|U|}$$

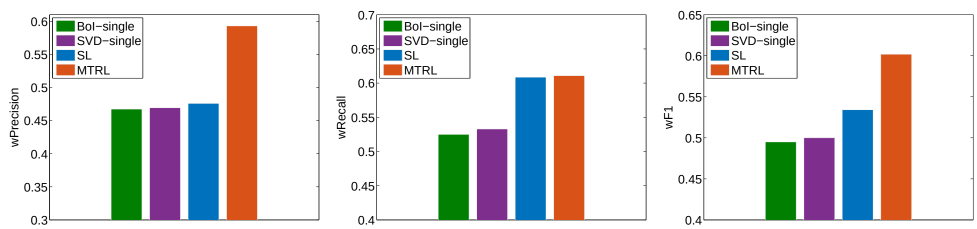
[2] http://www.brjt.cn/
[3] http://www.csie.ntu.edu.tw/∼cjlin/liblinear/
[4] http://tedlab.mit.edu/∼dr/SVDLIBC/



(a) gender



(b) marital status



(c) education background

Fig. 2: The performance comparison of different methods on BeiRen dataset.

$$\text{wRecall} = \frac{1}{|U|}\sum_{i} I(\mathcal{Y}^t_{(i)} = y^t_{(i)})$$

$$\text{wF1} = \frac{2 \times \text{wPrecision} \times \text{wRecall}}{\text{wPrecision} + \text{wRecall}}$$

where $\mathcal{Y}^t_{(i)}$ represents the calculated label of user i under task t, $I(\cdot)$ is an indicator function. Note here we use the weighted evaluation metrics because every class in task gender, marital status and education is as important as each other. As we can see, the weighted recall is the prediction accuracy in the user view.

The performance of different methods is shown in Figure 2. We have the following observations:

– (1) Using SVD to obtain low-dimension representations of users can achieve a better performance than BoI on predicting each demographic attribute of users. This result is quite accordance with the previous finds [23, 12].
– (2) Both the deep models SL and MTRL perform better than SVD-single. The result demonstrates that the deep model can learn a better representa-

Table 2: Performance comparison of different methods in terms of wF1 on BeiRen over different user groups.

user activeness	method	Gender	Martial Status	Education Background
unactive	BoI-single	0.522	0.660	0.403
	SVD-single	0.571	0.729	0.401
	SL	0.614	**0.747**	0.413
	MTRL	**0.678**	0.729	**0.415**
medium	BoI-single	0.589	0.686	0.506
	SVD-single	0.591	0.754	0.536
	SL	0.634	0.768	0.558
	MTRL	**0.645**	**0.802**	**0.647**
active	BoI-single	0.587	0.691	0.523
	SVD-single	0.568	**0.742**	0.526
	SL	0.646	0.716	0.533
	MTRL	**0.658**	0.732	**0.628**

tion comparing with the shallow one(here we regard SVD-single as a shallow model).

- (3) MTRL can improve the performance of each demographic prediction task significantly, especially to the education prediction task with limited data.
- (4) By using a multi-task approach to learn a shared representation layer across tasks, we can obtain a better performance than SL, which proves that the correlations among demographic attributes are helpful. MTRL can achieve the best performance in terms of all the evaluation measures, for example, when compared with the second best method(SL), the improvement of weighted F1-Measure on gender, marital status, and education background is 2.6%, 1.6%, and 6.4% respectively. By conducting the significant test, we find that the improvement of MTRL over the SL method is significant (p-$value < 0.01$) in terms of all the evaluation metrics.

To further investigate the performance of different methods, we split the users into three groups (i.e. inactive, medium and active) based on their activeness and conducted the comparisons on different user groups. We treat the user as inactive if there are less than 100 items in his/her purchase history, and active if there are more than 500 items in the purchase history. The remaining users are taken as medium. In this way, the proportions of inactive, medium and active are 45.1%, 42.9%, and 12.0% respectively. The results are shown in Table 2.

From the results we can see that, not surprisingly, the BoI-single method is still the worst on all the groups. Furthermore, by reducing user representations into a low dimensional space, SVD-single, SL and MTRL can performance better than BoI-single in all groups. Finally, MTRL can achieve the best performances on most groups in terms of all the measures. The results demonstrate that by learning a general representation using a multi-task approach, we can enjoy the relationships among demographic attributes, and complement each other to achieve a better performance.

5 Conclusion

In this paper, we try to predict users' demographic attributes given users' purchase behaviors . We propose a robust and practical representation learning algorithm MTRL based on multi-task objectives. Our MTRL can learn a shared representation across tasks, thus the sparseness problem can be avoided, especially for the task with limited data. Experiments on real-world purchase dataset demonstrate that our model can outperform the state-of-the-art baselines consistently under different evaluation metrics.

Although the MTRL model is proposed in this retail scenario, it is in fact a general model which can be applied on other multi-task multi-class problems. In the future, we would like to extend the usage of our MTRL model to model more demographic attributes to verify its effectiveness. Moreover, in this paper, we represent each user by simple bag of items as the raw input. It would be interesting to further explore the natural transaction structures in users' purchase data for a better demographic prediction.

6 Acknowledge

This research work was funded by 863 Program of China award number under Grant 2014AA015204, 973 Program of China award number under Grant 2014CB340401 ,2012CB316303, National Natural Science Foundation of China award numbers under Grant 61472401, 61433014, 61203298, 61425016, and Key Research Program of the Chinese Academy of Sciences under Grant NO.KGZD-EW-T03-2, and the Youth Innovation Promotion Association CAS under Grant no.20144310, and the Technology Innovation and Transformation Program of Shandong (Grant No.2014CGZH1103).

References

1. B. Bi, M. Shokouhi, M. Kosinski, and T. Graepel. Inferring the demographics of search users: Social data meets search queries. In *Proceedings of the 22Nd International Conference on World Wide Web*, pages 131–140, 2013.
2. C. J. Culotta A, Ravi N K. Predicting the demographics of twitter users from website traffic data[c]. *ICWSM, in press. Menlo Park, California: AAAI Press*, 2015.
3. Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla. Inferring user demographics and social strategies in mobile social networks. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 15–24, 2014.
4. S. Duarte Torres and I. Weber. What and how children search on the web. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 393–402, New York, NY, USA, 2011. ACM.
5. P. Eckert. Gender and sociolinguistic variation. *Readings in Language and Gender*, 1997.

6. T. Evgeniou and M. Pontil. Regularized multi–task learning. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 109–117. ACM, 2004.
7. H. G. Graves A, Mohamed A. Speech recognition with deep recurrent neural networks. *Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013.
8. J. Hu, H.-J. Zeng, H. Li, C. Niu, and Z. Chen. Demographic prediction based on user's browsing behavior. In *Proceedings of the 16th International Conference on World Wide Web*, pages 151–160. ACM, 2007.
9. P. D. S. Kalyanam K. Incorporating demographic variables in brand choice models: An indivisible alternatives framework. *Marketing Science*, 1997.
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, page 2012.
11. X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang. Representation learning using multi-task deep neural networks for semantic classification and information retrieval. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015.
12. D. S. M. Kosinski and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 2013.
13. C. Micchelli and M. Pontil. Kernels for multi-task learning. *NIPS*, 2005.
14. A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You are who you know: Inferring user profiles in online social networks. In *WSDM*, pages 251–260, 2010.
15. A. Mnih and G. Hinton. Three new graphical models for statistical language modelling. In *Proceedings of the 24th International Conference on Machine Learning*, pages 641–648, 2007.
16. D. Murray and K. Durrell. Inferring demographic attributes of anonymous internet users. *In Web Usage Analysis and User Profiling*, 2000.
17. J. Otterbacher. Inferring gender of movie reviewers: Exploiting writing style, content and metadata. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 369–378, New York, NY, USA, 2010. ACM.
18. I. S. C. Rick L. Andrews. Identifying segments with identical choice behaviors across product categories: An intercategory logit mixture model. *International Journal of Research in Marketing*, 2002.
19. J. Schler, M. Koppel, S. Argamon, and J. W. Pennebaker. Effects of age and gender on blogging. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pages 199–205. AAAI, 2006.
20. S. Sedhain, S. Sanner, D. Braziunas, L. Xie, and J. Christensen. Social collaborative filtering for cold-start recommendations. In *Proceedings of the 8th ACM Conference on Recommender Systems*, pages 345–348, 2014.
21. S. S. You Ji. Multitask multiclass support vector machines. *Data Mining Workshops (ICDMW)*, 2011.
22. E. Zhong, B. Tan, K. Mo, and Q. Yang. User demographics prediction based on mobile data. *Pervasive Mob. Comput.*, 9(6):823–837, Dec. 2013.
23. Y. Zhong, N. J. Yuan, W. Zhong, F. Zhang, and X. Xie. You are where you go: Inferring demographic attributes from location check-ins. In *WSDM*, WSDM '15, pages 295–304, New York, NY, USA, 2015. ACM.